A common error that learners run into with the week 3 assignment is looking at the error bars of the data rather

than the error bars of the means of the data. These are very different, as the standard deviation of the means involves

taking the square root of the number of samples.

This reading is intended to clarify the process required for assignment 3, with the demonstration based on the 1992 portion of

the following data; we will create 1000 samples with a set random seed for reproducibility.

```
import pandas as pd

import numpy as np

df = pd.DataFrame([np.random.normal(32000,200000,3650),

                   np.random.normal(43000,100000,3650),

                   np.random.normal(43500,140000,3650),

                   np.random.normal(48000,70000,3650)],

                  index=[1992,1993,1994,1995])


# Let's do the random sampling 1000 times

np.random.seed(12345)

df_means = pd.DataFrame({'means':[np.random.normal(32000,200000,3650).mean() for i in range(1000)]})

df_means.head()
```

#means head ouput:

033312.107476

129723.719082

226276.149916

331267.288484

431121.673831

Using the 1000 samples of means, we will now compute the standard deviation.

df_means.std(axis=0)

#std output:

means 3414.816232

dtype: float64

This standard deviation is that of the means (also known as the standard error), and is the standard deviation

used for the error bars. Note that this is not the standard deviation of the data.

The formula for calculating standard error is as follows (see this Wikipedia article for more):

Sx= S / (Raiz cuadrada (n))

Using the above formula, we can calculate the standard error as follows:

data standard deviation: 200000

sample size: 3650

import math

200000 / math.sqrt(3650)

#output:

3310.4235544094718